



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

## NOTICE OF ALLOWANCE AND FEE(S) DUE

20529 7590 03/04/2010

THE NATH LAW GROUP  
112 South West Street  
Alexandria, VA 22314

EXAMINER

RUIZ, ANGELICA

ART UNIT

PAPER NUMBER

2158

DATE MAILED: 03/04/2010

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/573,482

03/24/2006

David Patterson

27309U

2569

TITLE OF INVENTION: COMPUTER AIDED DOCUMENT RETRIEVAL

APPLN. TYPE	SMALL ENTITY	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	YES	\$755	\$300	\$0	\$1055	06/04/2010

**THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.**

**THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.**

### HOW TO REPLY TO THIS NOTICE:

#### I. Review the SMALL ENTITY status shown above.

If the SMALL ENTITY is shown as YES, verify your current SMALL ENTITY status:

A. If the status is the same, pay the TOTAL FEE(S) DUE shown above.

B. If the status above is to be removed, check box 5b on Part B - Fee(s) Transmittal and pay the PUBLICATION FEE (if required) and twice the amount of the ISSUE FEE shown above, or

If the SMALL ENTITY is shown as NO:

A. Pay TOTAL FEE(S) DUE shown above, or

B. If applicant claimed SMALL ENTITY status before, or is now claiming SMALL ENTITY status, check box 5a on Part B - Fee(s) Transmittal and pay the PUBLICATION FEE (if required) and 1/2 the ISSUE FEE shown above.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

**IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.**

# **PART B - FEE(S) TRANSMITTAL**

**Complete and send this form, together with applicable fee(s), to: Mail Mail Stop ISSUE FEE  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
or Fax (571)-273-2885**

**INSTRUCTIONS:** This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

20529 7590 03/04/2010

**THE NATH LAW GROUP**  
112 South West Street  
Alexandria, VA 22314

**Note:** A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

## **Certificate of Mailing or Transmission**

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/573,482	03/24/2006	David Patterson	27309U	2569

**TITLE OF INVENTION:** COMPUTER AIDED DOCUMENT RETRIEVAL

APPLN. TYPE	SMALL ENTITY	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	YES	\$755	\$300	\$0	\$1055	06/04/2010

EXAMINER	ART UNIT	CLASS-SUBCLASS
RUIZ, ANGELICA	2158	707-005000

**1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).**

- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
- ☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

**2. For printing on the patent front page, list**

- (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, 1 \_\_\_\_\_
- (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed. 2 \_\_\_\_\_
- 3 \_\_\_\_\_

**3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)**

**PLEASE NOTE:** Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☐ Corporation or other private group entity ☐ Government

**4a. The following fee(s) are submitted:**

- ☐ Issue Fee
- ☐ Publication Fee (No small entity discount permitted)
- ☐ Advance Order - # of Copies \_\_\_\_\_

**4b. Payment of Fee(s): (Please first reapply any previously paid issue fee shown above)**

- ☐ A check is enclosed.
- ☐ Payment by credit card. Form PTO-2038 is attached.
- ☐ The Director is hereby authorized to charge the required fee(s), any deficiency, or credit any overpayment, to Deposit Account Number \_\_\_\_\_ (enclose an extra copy of this form).

**5. Change in Entity Status (from status indicated above)**

- ☐ a. Applicant claims SMALL ENTITY status. See 37 CFR 1.27. ☐ b. Applicant is no longer claiming SMALL ENTITY status. See 37 CFR 1.27(g)(2).

**NOTE:** The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

Authorized Signature \_\_\_\_\_

Date \_\_\_\_\_

Typed or printed name \_\_\_\_\_

Registration No. \_\_\_\_\_

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. **DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.**

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/573,482	03/24/2006	David Patterson	27309U	2569
20529	7590	03/04/2010	EXAMINER	
THE NATH LAW GROUP 112 South West Street Alexandria, VA 22314			RUIZ, ANGELICA	
			ART UNIT	PAPER NUMBER
			2158	
DATE MAILED: 03/04/2010				

## Determination of Patent Term Adjustment under 35 U.S.C. 154 (b) (application filed on or after May 29, 2000)

The Patent Term Adjustment to date is 44 day(s). If the issue fee is paid on the date that is three months after the mailing date of this notice and the patent issues on the Tuesday before the date that is 28 weeks (six and a half months) after the mailing date of this notice, the Patent Term Adjustment will be 44 day(s).

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (<http://pair.uspto.gov>).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

<b>Notice of Allowability</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	10/573,482	PATTERSON ET AL.	
	<b>Examiner</b>	<b>Art Unit</b>	
	ANGELICA RUIZ	2158	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--**

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to 12/14/2009.
2. ☒ The allowed claim(s) is/are 2-12, 14-17, an d19-21 (Renumbered 1-18).
3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a) ☐ All    b) ☐ Some\*    c) ☐ None    of the:
    1. ☐ Certified copies of the priority documents have been received.
    2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

\* Certified copies not received: \_\_\_\_\_.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.  
**THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.**

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
5. ☐ CORRECTED DRAWINGS ( as "replacement sheets") must be submitted.
  - (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review ( PTO-948) attached
    - 1) ☐ hereto or 2) ☐ to Paper No./Mail Date \_\_\_\_\_.
  - (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date \_\_\_\_\_.

**Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).**
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

**Attachment(s)**

- |  |   |
|--|---|
| <ol style="list-style-type: none"> <li>1. <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)</li> <li>2. <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)</li> <li>3. <input type="checkbox"/> Information Disclosure Statements (PTO/SB/08),<br/>Paper No./Mail Date _____</li> <li>4. <input type="checkbox"/> Examiner's Comment Regarding Requirement for Deposit<br/>of Biological Material</li> </ol> | <ol style="list-style-type: none"> <li>5. <input type="checkbox"/> Notice of Informal Patent Application</li> <li>6. <input type="checkbox"/> Interview Summary (PTO-413),<br/>Paper No./Mail Date _____.</li> <li>7. <input checked="" type="checkbox"/> Examiner's Amendment/Comment</li> <li>8. <input checked="" type="checkbox"/> Examiner's Statement of Reasons for Allowance</li> <li>9. <input type="checkbox"/> Other _____.</li> </ol> |
|--|---|

/Angelica Ruiz/  
Examiner, Art Unit 2158

/Mohammad Ali/  
Supervisory Patent Examiner, Art Unit 2158

### **DETAILED ACTION**

1. This office action is in response to applicant's response to non final office action filed on April 30, 2009 and subsequent supplemental on December 14, 2009 amendment on in which claims 1-12, 14-17, and 19-21 are presented for further examination.

In response of Applicants amendments and remarks arguments with respect to the rejection of claims 1-12, 14-18 under Choi (US 2002/0042793 A1) in view of Deerwester et al (US 4,839,853) have been fully considered. After the proposed changes presented in the present Examiner's amendment, the rejection has been withdrawn.

### **EXAMINER'S AMENDMENT**

2. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it **MUST** be submitted no later than the payment of the issue fee.

3. Authorization for this examiner's amendment was given in a telephone interview with Jiaxiao Zhang on February 26, 2010.

The application has been amended as follows:

**In the claim:**

See Attached **Exhibit A**.

***Reason for Allowance***

4. The following is an examiner's statement of reasons for allowance: The present invention relates to computer systems and particularly to replication of application checkpoint data. The closest prior art Choi (US 2002/0042793 A1) is directed to a method of order-ranking document clusters using entropy data and Bayesian self-organizing feature maps. Deerwester also is directed to Computer information retrieval using latent semantic structure. However, Choi and Deerwester either singularly or in combination, fail to anticipate or render obvious the recited features "A computer-implemented method of determining cluster attractors for use in clustering a plurality of documents, each document comprising at least one term, each term comprising one or more words, the method comprising: causing a computer to calculate, in respect of each term, a probability distribution that is indicative of in the instance where a document comprises said term and said one other term that co-occurs with said term in at least one of said documents, the frequency of occurrence of said one other term in the instance where a document comprises said term and said one other term, and in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, the respective frequency of occurrence of each other term, that co-occurs with said term in at least one of said documents; causing a computer to calculate, in respect of each term, the entropy of the respective probability distribution; and causing the computer to select at least one of said probability distributions as a cluster attractor depending on the

Art Unit: 2158

respective entropy value, wherein the selected cluster attractor is a clustering focus for at least some of said documents, and wherein said probability distribution is calculated

as

$$p(y|z) = \frac{\sum_{x \in X(z)} tf(x,y)}{\sum_{x \in X(z)} tf(x,t)}$$

where  $tf(x, y)$  is a term frequency of a term  $y$  in a document  $x$  and  $X(z)$  is a set of all documents of said plurality of documents that contain a term  $z$  and where  $t$  is a term index". These features in conjunction with all other limitations of the dependents and independent claims render claims 2-12, 14-17, 19-21 (Renumbered 1-18) allowable.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

### **Conclusion**

5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to ANGELICA RUIZ whose telephone number is (571)270-3158. The examiner can normally be reached on 8:00 a.m. to 4:30 p.m., ET.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Mohammad Ali can be reached on (571) 272-4105. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2158

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Angelica Ruiz/  
Examiner, Art Unit 2158

/Mohammad Ali/  
Supervisory Patent Examiner, Art Unit 2158



Exhibit A:

**Listing of Claims:**

Claim 1. (Canceled)

Claim 2. (Currently amended)      A method as claimed in ~~Claim 1~~ claim 20, wherein each probability distribution comprises, in respect of each co-occurring term, an indicator that is indicative of the total number of instances of the respective co-occurring term in all of the documents in which the respective co-occurring term co-occurs with the term in respect of which the probability distribution is calculated.

Claim 3. (Currently amended)      A method as claimed in ~~Claim 1~~ claim 20, wherein each probability distribution comprises, in respect of each co-occurring term, an indicator comprising a conditional probability of the occurrence of the respective co-occurring term in a document given the appearance in said document of the term in respect of which the probability distribution is calculated.

Claim 4. (Previously presented)      A method as claimed in Claim 2, wherein each indicator is normalized with respect to the total number of terms in the document, or each document in which the term in respect of which the probability distribution is calculated appears.

Art Unit: 2158

Claim 5. (Currently amended) A method as claimed in ~~Claim 4~~ claim 20, comprising assigning each term to one of a plurality of subsets of terms depending on the frequency of occurrence of the term; and selecting, as a cluster attractor, the respective probability distribution of one or more terms from each subset of terms.

Claim 6. (Original) A method as claimed in Claim 5, wherein each term is assigned to a subset depending on the number documents of the corpus in which the respective term appears.

Claim 7. (Previously presented) A method as claimed in Claim 5, wherein an entropy threshold is assigned to each subset, the method comprising selecting, as a cluster attractor, the respective probability distribution of one or more terms from each subset having an entropy that satisfies the respective entropy threshold.

Claim 8. (Original) A method as claimed in Claim 7, comprising selecting, as a cluster attractor, the respective probability distribution of one or more terms from each subset having an entropy that is less than or equal to the respective entropy threshold.

Claim 9. (Previously presented) A method as claimed in Claim 5, wherein each subset is associated with a frequency range and wherein the frequency ranges for respective subsets are disjoint.

Art Unit: 2158

Claim 10. (Previously presented) A method as claimed in Claim 5, wherein each subset is associated with a frequency range, the size of each successive frequency range being equal to a constant multiplied by the size of the preceding frequency range in order of increasing frequency.

Claim 11. (Previously presented) A method as claimed in Claim 7, wherein the respective entropy threshold increases for successive subsets in order of increasing frequency.

Claim 12. (Original) A method as claimed in Claim 11, wherein the respective entropy threshold for successive subsets increases linearly.

Claim 13. (Canceled)

Claim 14. (Currently amended) An apparatus for determining cluster attractors for a plurality of documents, each document comprising at least one term, each term comprising one or more words, the apparatus comprising:

means for calculating, in respect of each term, a probability distribution indicative of

in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents, the frequency of occurrence of said one other term, and

in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, the respective frequency of

occurrence of each other term;

means for calculating, in respect of each term, the entropy of the respective probability distribution; and

means for selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value,

wherein the selected cluster attractor is a clustering focus for at least some of said documents, and wherein said probability distribution is calculated as

$$p(y|z) = \frac{\sum_{x \in X(z)} f(x,y)}{\sum_{x \in X(z)} f(x,z)}$$

where  $f(x,y)$  is a term frequency of a term  $y$  in a document  $x$  and  $X(z)$  is a set of all

documents of said plurality of documents that contain a term  $z$  and where  $i$  is a term index.

Claim 15. (Currently amended) ~~A computer-implemented method of clustering a plurality of documents; each document comprising at least one term; each term comprising one or more words; the method comprising;~~

~~causing a computer to calculate, in respect of each term, a probability distribution indicative of~~

~~in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents; the frequency of occurrence of said one other term; and~~

~~in the instance where a document comprises said term and more than one other term~~

~~that co-occurs with said term in at least one of said documents, the respective frequency of occurrence of each other term;~~

~~The method as claimed in claim 20, wherein the method further comprises:~~

~~causing a computer to calculate, in respect of each term, the entropy of the respective probability distribution;~~

~~causing the computer to select at least one of said probability distributions as a cluster attractor depending on the respective entropy value;~~

~~causing the computer to compare each document with each cluster attractor; and~~

~~causing the computer to assign each document to one or more cluster attractors depending on the similarity between the document and the cluster attractors,~~

~~wherein assigning each document to one or more cluster attractors creates a plurality of document clusters, each cluster comprising a respective plurality of documents.~~

Claim 16. (Original) A method as claimed in Claim 15, comprising: calculating, in respect of each document, a probability distribution indicative of the frequency of occurrence of each term in the document; comparing the respective probability distribution of each document with each probability distribution selected as a cluster attractor; and assigning each document to at least one cluster depending on the similarity between the compared probability distributions.

Claim 17. (Previously presented) A method as claimed in Claim 16, comprising organizing the documents within each cluster by: assigning a respective weight to each document, the value of the

Art Unit: 2158

weight depending on the similarity between the probability distribution of the document and the probability distribution of the cluster attractor; comparing the respective probability distribution of each document in the cluster with the probability distribution of each other document in the cluster; assigning a respective weight to each pair of compared documents, the value of the weight depending on the similarity between the compared respective probability distributions of each document of the pair; calculating a minimum spanning tree for the cluster based on the respective calculated weights.

Claim 18. (Canceled)

Claim 19. (Currently amended) ~~An apparatus for clustering a plurality of documents, each document comprising at least one term, each term comprising one or more words, the apparatus comprising:~~

~~means for calculating, in respect of each term, a probability distribution indicative of~~  
~~in the instances where a document comprises said term and one other term that co-occurs~~  
~~with said term in at least one of said documents, the frequency of occurrence of said one~~  
~~other term; and~~

~~in the instance where a document comprises said term and more than one other term~~  
~~that co-occurs with said term in at least one of said documents, the respective frequency of~~  
~~occurrence of each other term;~~

~~means for calculating, in respect of each term, the entropy of the respective probability~~  
~~distribution;~~

~~means for selecting at least one of said probability distributions as a cluster attractor  
depending on the respective entropy value;~~

~~The apparatus as claimed in claim 14, wherein the apparatus further comprises:~~

~~means for comparing each document with each cluster attractor; and~~

~~means for assigning each document to one or more cluster attractors depending on the  
similarity between the document and the cluster attractors,~~

~~wherein assigning each document to one or more cluster attractors creates a plurality of  
document clusters, each cluster comprising a respective plurality of documents.~~

20. (Currently amended) A computer-implemented method of determining cluster attractors for  
use in clustering a plurality of documents, each document comprising at least one term, each term  
comprising one or more words, the method comprising:

causing a computer to calculate, in respect of each term, a probability distribution that is  
indicative of

in the instance where a document comprises said term and said one other term that co-  
occurs with said term in at least one of said documents, the frequency of occurrence of said  
one other term in the instance where a document comprises said term and said one other  
term, and

in the instance where a document comprises said term and more than one other term  
that co-occurs with said term in at least one of said documents, the respective frequency of

Art Unit: 2158

occurrence of each other term, that co-occurs with said term in at least one of said documents;

causing a computer to calculate, in respect of each term, the entropy of the respective probability distribution; and

causing the computer to select at least one of said probability distributions as a cluster attractor depending on the respective entropy value,

wherein the selected cluster attractor is a clustering focus for at least some of said documents, and wherein said probability distribution is calculated as

$$p(y | z) = \frac{\sum_{x \in X(z)} f(x, y)}{\sum_{x \in X(z), t \in T} f(x, t)}$$

where  $f(x, y)$  is a term frequency of a term  $y$  in a document  $x$  and  $X(z)$  is a set of all documents of said plurality of documents that contain a term  $z$  and where  $t$  is a term index.

Claim 21. (Previously presented) A method as claimed in Claim 20, wherein said entropy is calculated as:

$$H(Y | z) = - \sum_y p(y | z) \log p(y | z)$$

/Angelica Ruiz/  
Examiner, Art Unit 2158